Chapter 5

Exocentric Sensing & Delivery:

Facial Expressions

As described in Chapter 2, in the survey conducted towards understanding the non-verbal cue needs for people who are blind and visually impaired, they emphasized on the lack of access to facial expressions and mannerisms of their interaction partners. This is supported by the argument that most part of the non-verbal cues occur through visual facial mannerisms as described in Section 1.2.1 of Chapter 1. The face encodes a lot of information that is both communicative and expressive in nature. Unfortunately, the face is a very complex data generator and the encodings on the face are not very context sensitive and individualistic in nature. Evolving computing technologies have been focused on developing solutions towards understanding the nature of facial mannerisms and gestures, but most of this multi-modal affective interaction research has been focused on the development of sensors and algorithms that understand user's emotional state in a human-machine interaction scenario. These interactions are mostly unilateral in nature and directed primarily towards the machine interpreting the user's emotional state. That is, the machines become the primary consumers of the affective cues. But from the perspective of an assistive technology affect interactions have to be augmentations that enrich human-human interpersonal interaction, where the machines not only interpret communicator’s affective state, but also delivers affect information through novel affect actuators to a social interaction recipient.

As mentioned before most affect information is causal in nature and understanding what the expression or mannerism means requires an understanding of context when it is happening and the situation in which the communication is occurring. Our understanding of the cognitive models within the human brain that allows for the processing of complex facial expressions and emotions is very naïve. Computational models developed towards understanding context are very simplistic and performs nominally even under very well controlled laboratory conditions. Contrary to such a setting, assistive technologies provide some respite to the complexities by having the cognitive abilities of the user of the technology to make decisions. That is, while human computer interfaces need to mimic sensing, cognition and delivery, assistive technologies for people who are blind have to look at sensing and delivery alone and piggy back on human cognition. This requires precise sensing of the facial and head movements while delivering as much information back to the user as possible through technologies that do not overload the user with information but provides just the right level of information to allow them to cognitively process this information.

Thus, the focus of this chapter is on the *precise sensing* and *proficient delivery* of facial mannerisms and gestures of interaction partners to the user of the Social Interaction Assistant who is blind or visually impaired. To this end, the two important aspects of sense and delivery will be handled simultaneously to meet the goal of delivering dynamic facial and head movement information to the user of the social interaction assistant.

*From the sensing perspective, current ongoing experiments in tracking of facial expressions and mannerisms will be described in detail with identified areas that need special attention.*

*From the delivery perspective, the latest in haptic interface will be introduced as a means of conveying facial and head mannerisms. Details on the experiments that have been carried out and the ones that need to be conducted will be illustrated.*

# Sensing Facial Mannerisms and Expressions:

## Head Movements:

### Real-time tracking of the head:

## Facial Feature Movements:

### Registration of the head and face:

#### Image features for expression recognition:

#### Facial feature movement for dynamic motion conveyance:

## Experiments:

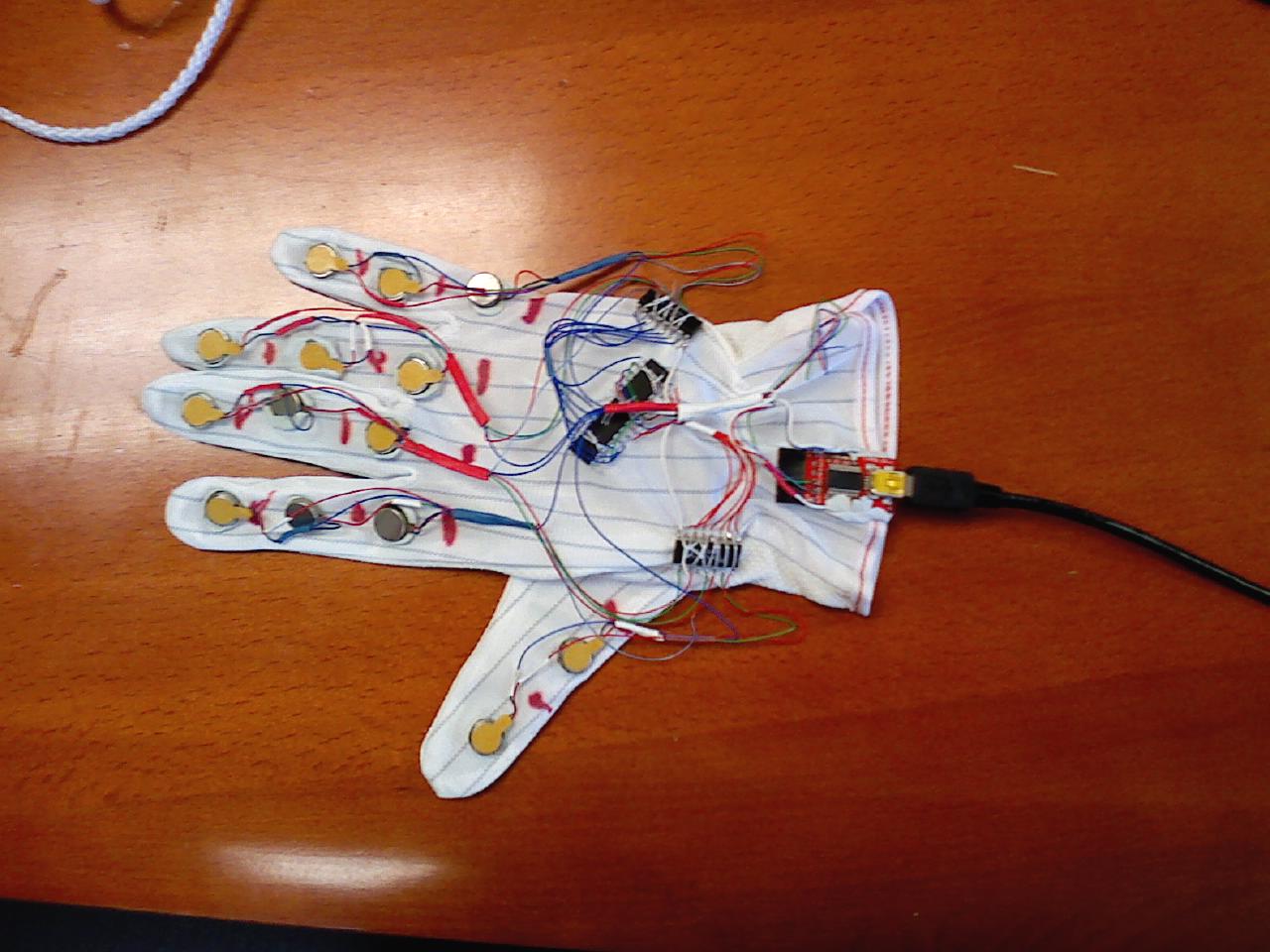
# Delivering Facial Mannerisms and Expressions:

## Design Considerations:

People who are blind rely on their auditory senses to understand and comprehend the environment around them. As described in detail in Chapter 1 Section 1.5.1, assistive technologies that use audio cues to deliver information back to a user can cause sensory overload leading to the rejection of any benefits that a device might offer. Especially during social interactions and bilateral conversations, it is imperative that any device should not hinder the primary sensory channel of the user. As shown in Section 1.7.1 of Chapter 1, Haptics offers a high-bandwidth channel for delivering information. As seen from the human homunculus, the hands form a perfect region to deliver this high bandwidth data. Of the various dimensions of somatosensory perception of the human skin, we choose to work with vibrators that can actuate the Meissner’s Corpuscles or the Pacinian Corpuscles thereby allowing amplitude, frequency and rhythm as the primary dimensions to work with. A detailed background work on the use of vibrotactile actuations to convey information through the human hands can be found in Section 1.7.4 of Chapter 1. Here we describe in detail the construction of the vobrotactile glove and the mappings used for conveying facial expressions.

## Construction of the Haptic Glove:

The proposed vibrotactile glove was built The haptic glove has 14 tactors (vibration motors) mounted on the back of the fingers, one per phalange. The 14 motors correspond to the 14 phalanges (3 each on the index finger, middle finger, ring finger and the pinky with 2 on the thumb) on the human hand. A controller is also integrated on the glove to allow control of the motor’s vibration (magnitude, duration and temporal rhythm) through the USB port of a PC.



**Figure 1:** Haptic Glove: The figure shows a glove made out of stretchable material with 14 motors on the back of the glove with each motor corresponding to one phalange of the 5 digits. A microcontroller, two motor drivers and 1 USB controller (4 ICs) are also integrated on to the back of the glove with an ultra thin flexible USB cable leaving the glove.

Currently, the device only delivers the 6 basic expressions (Smile, Anger, Disgust, Surprise, Sad and Fear) along with indications of when the face reaches neutral expression. In future, we plan to encode the dynamic motion of the human facial features into vibrotactile patterns. This would allow indiscriminate access to the facial movements of the interaction counterpart. Due to the lack of space we do not discuss the details of the design process, but we introduce, in brief, the construction of the haptic device, the vibotactile structure of the expressions and describe an experiment carried out on the haptic glove.

## Mapping for facial expressions:

In order to encode the 6 basic expressions and neutral facial posture into haptic cues, we resorted to popular emoticon representations of these basic expressions. For example, smile is popularly represented by a smiley which was translated to a vibratory pattern of index finger top phalange, followed by middle finger bottom, followed by ring finger top phalange. The entire vibration sequence was completed within 750 milliseconds (The duration was arrived at after careful pilot studies with participants). The table below gives the vibration finger and phalange location in comma separated sequence for all 7 facial expression postures.

|  |  |
| --- | --- |
| Expression | Comma separated vibration sequence. All sequences are 750ms long  First letter indicates the finger – I for index, M for middle and R for Ring  Second letter indicates the phalange – T for top, M for middle and B for bottom |
| Smile | IT, MB, RT |
| Sad | IB, MT, RB |
| Surprise | MT, IM, MB, RM, MT |
| Anger | IM, IB, MM, MB, RM, RB |
| Neutral | IM, MM, RM |
| Disgust | RB, MB, IB |
| Fear | IT, MT, RT, MT, IT, MT, RT |

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Happy | |  |  |  | | --- | --- | --- | | • |  | • | |  |  |  | |  | • |  | |  | expressions.jpg |
| Sad | |  |  |  | | --- | --- | --- | |  | • |  | |  |  |  | | • |  | • | |  | expressions.jpg |
| Surprise | |  |  |  | | --- | --- | --- | |  | • |  | | • |  | • | |  | • |  | |  | expressions.jpg |
| Neutral | |  |  |  | | --- | --- | --- | |  |  |  | | • | • | • | |  |  |  | |  |  |
| Angry | |  |  |  | | --- | --- | --- | |  |  |  | | • | • | • | | • | • | • | |  | 6a00d8345202e469e200e54f652e998833-800wi.jpg |
| Fear | |  |  |  | | --- | --- | --- | | • | • | • | |  |  |  | |  |  |  | |  | expressions.jpg |
| Disgust | |  |  |  | | --- | --- | --- | |  |  |  | |  |  |  | | • | • | • | |  | expressions.jpg |

## Experiments:

The above expressions were conveyed to 10 participants one of whom is blind. The participants were trained on the expressions until they were able to recognize all the expressions without any mistake after which 70 stimulations (10 trails of each expression) were presented sequentially with 5 seconds gap between each for the user to respond. The table below represents the results as a 7x7 confusion matrix where each cell entry corresponds to how many times (on average) users when given the row expression as stimulation responded with the column expression as their answer. Following this average number, separated by a comma is the average time taken for answering.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Response** | | | | | | | |
| **Stimulation** |  | **Angry** | **Disgust** | **Fear** | **Smile** | **Sad** | **Surprise** | **Neutral** |
| **Angry** | 88, 2.28 | - | - | 1, 4.45 | - | 11, 2.99 | - |
| **Disgust** | - | 95, 1.89 | 2, 1.52 | - | 2, 3.44 | - | 1, 1.67 |
| **Fear** | - | 1, 2.64 | 98, 1.66 | - | - | 1, 3.62 | - |
| **Smile** | - | - | - | 88, 2.12 | 5, 4.18 | 7, 2.8 | - |
| **Sad** | 4, 2.54 | - | - | 2, 2.32 | 82, 2.67 | 10, 2.68 | 2, 2.38 |
| **Surprise** | 10, 3.12 | - | 2, 2.71 | - | 1, 3.48 | 87, 2.52 | - |
| **Neutral** | 2, 2.69 | - | 1, 2.94 | 3, 3.6 | 8, 3.56 | - | 86, 2.23 |
|  |  |  |  |  |  |  |  |  |

**Table 1:** Rows represent the stimulation provided to the users and the columns represent the response provided by the user. Each cell has two numbers. The first number represents the percentage recognition of a specific stimulation and a corresponding response. The second number represents the average time taken in seconds for that specific stimulation and response. Ideally this matrix should have 100% recognition along the diagonal with as low a time as possible.